6/PRB 09/743745
500 [redacted]'d PCT/PTO 16 JAN 2001

SILICON-CONTAINING LINKERS FOR NUCLEIC ACID MASS MARKERS

This invention concerns compounds which comprise mass markers for detection by mass spectrometry. The invention relates to methods for characterising nucleic acids or other molecules by labelling with markers that are cleavably detachable from their associated nucleic acid and that are detectable by mass spectrometry. Specifically this invention relates to improved methods of detaching mass labels from their associated nucleic acids or other molecules of interest.

PCT/GB98/00127 describes arrays of cleavable labels that are detectable by mass spectrometry which identify the sequence of a covalently linked nucleic acid probe. These mass labels have a number of advantages over other methods of analysing nucleic acids. At present commercially favoured systems are based on fluorescent labelling of DNA. Fluorescent labelling schemes permit the labelling of a relatively small number of molecules simultaneously, typically 4 labels can be used simultaneously and possibly up to eight. However the costs of the detection apparatus and the difficulties of analysing the resultant signals limit the number of labels that can be used simultaneously in a fluorescence detection scheme. An advantage of using mass labels is the possibility of generating large numbers of labels which have discrete peaks in a mass spectrum allowing similar numbers of distinct molecular species to be labelled simultaneously. Fluorescent dyes are expensive to synthesise whereas mass labels can comprise relatively simple polymers permitting combinatorial synthesis of large numbers of labels at low cost.

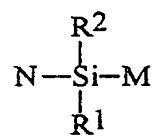
A feature of the mass labelling techniques disclosed in PCT/GB98/00127 is the need for linker groups that covalently link a mass marker to its corresponding nucleic acid. These linkers must permit the mass marker to be separated from its nucleic acid prior to detection within a mass spectrometer. It is desirable that the cleavage of the label from its nucleic acid be performed in-line with a mass spectrometer, possibly after some in-line pre-fractionation step such as capillary electrophoresis. It is also desirable that this in-line cleavage step does not require a complex interface with the mass spectrometer to enable this step to occur. Ideally

linkers should cleave at some predetermined point within existing instruments without any modification to the instrument beyond changes of normal operating parameters.

Linkers should cleave without damaging associated nucleic acids hence reducing noise in the mass spectrum from nucleic acid fragmentation. Linkers should all cleave under the same conditions to ensure all labels can be analysed simultaneously and quantitatively.

It is an object of this invention to provide linkers that have the desired features disclosed above which are compatible with existing mass spectrometers particularly electrospray ionisation and tandem mass spectrometry.

Accordingly, the present invention provides a compound having the following formula (I):



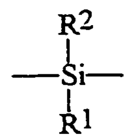
(I)

wherein M comprises a mass marker, N comprises a nucleic acid and R^1 and R^2 are substituents selected such that when the compound reacts with an electron donating moiety, either N or M cleaves from the Si atom in preference to R^1 and R^2.

Compounds with the formula (I) shown above meet the specification discussed above. The molecule is stable during synthesis and can be cleaved under mild conditions in an electrospray ion source or in the collision chamber of a tandem mass spectrometer in the presence of an appropriately reactive gaseous electron donating moiety, such as ammonia. The reactive gas participates in a novel gas phase reaction with the linker resulting in the cleavage of the linker.

The present invention also provides a method for characterising an analyte, which method comprises:

(a) providing a compound in which the analyte is attached by a cleavable linker to a reporter group relatable to the analyte, the linker having the following formula:

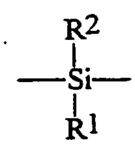


wherein R^1 and R^2 are substituents as defined below;

(b) cleaving the reporter group from the analyte; and

(c) identifying the reporter group, thereby characterising the analyte.

The invention additionally provides use of a linker group in the characterisation of an analyte, to attach a reporter group to the analyte, wherein the linker group is cleavable and has the following formula:



wherein R^1 and R^2 are substituents as defined below.

The invention will now be described in further detail by way of example only, with reference to the accompanying drawings, in which:

Figure 1 depicts the mechanism of cleavage of a linker used in the present invention, by means of a primary amine. Cleavage takes place via a five-co-ordinate intermediate to produce two possible products. The mass spectrum of the charged products is measured;

Figure 2 shows the negative ion mass spectrum of FT23 at 0.35 pmol/μl. This spectrum shows a very clean molecular ion less 1 proton at m/z 729.4;

Figure 3 shows the positive ion mass spectrum of FT23 at 0.35 pmol/μl. This spectrum shows a strong protonated molecular ion peak at m/z 731.5 and a second stronger peak at m/z 753.3 corresponding to a sodium adduct of FT23;

Figure 4 shows the negative ion mass spectrum of FT23 at 0.35 pmol/μl in the presence of a PCR product. The molecular ion is not detectable over the background peaks;

Figure 5 shows the positive ion mass spectrum of FT23 at 0.35 pmol/μl in the presence of a PCR product. This spectrum does not show any of the protonated molecular ion peak at m/z 731.4 or the sodium adduct peak at m/z 753.3. Two additional peaks of significance appear in this spectrum. One peak at m/z 739.2 and a second peak at m/z 755.2. These peaks are believed to be end-products of a novel gas phase cleavage reaction discussed below. The two peaks corresponding to gas phase cleavage products are the only major ion peaks in this spectrum; and

Figure 6 shows a reaction mechanism for ammonia reacting with a TBDMS protective group used in the present invention.

In the methods of the present invention, the analyte is not particularly limited and can be any analyte or molecule of interest, such as a nucleic acid or other molecule. Typically the analyte comprises a biological molecule. In preferred embodiments of the present invention, the biological molecule is selected from a protein, a polypeptide, an amino acid, a nucleic acid (e.g. an RNA, a DNA, a plasmid, a nucleotide or an oligonucleotide), a nucleic acid base, a pharmaceutical agent or drug, a carbohydrate, a lipid, a natural product and a synthetic compound from an encoded chemical library. When the analyte comprises a nucleotide, oligonucleotide or nucleic acid, the nucleotide, oligonucleotide or nucleic acid may be natural,

or may be modified by modifying a base, sugar and/or backbone of the nucleotide, oligonucleotide or nucleic acid. In the compounds of the present invention, the analyte is a nucleic acid, and may be any type of nucleic acid. Preferably, the nucleic acid is of a type as defined above.

The substituents R^1 and R^2 are not especially limited. It is preferred that R^1 and R^2 are selected such that their bond energies to Si are greater than the bond energy of N and/or M to Si to ensure that when the compound is reacted with an electron donating moiety either N or M cleaves from the Si atom in preference to R^1 and R^2, and/or R^1 and R^2 are selected such that their steric bulk is sufficient to ensure that when the compound is reacted with an electron donating moiety either N or M cleaves from the Si atom in preference to R^1 and R^2. Typically, R^1 and R^2 are each independently a hydrogen atom, a halogen atom, a substituted or unsubstituted alkyl group, or a substituted or unsubstituted aryl group. It is particularly preferred that R^1 and R^2 are each independently fluorine, chlorine, bromine, iodine, methyl, ethyl, propyl, isopropyl, butyl, isobutyl, tert-butyl or phenyl groups.

Thus, various substituents may be introduced at the positions R^1 and R^2 including fluorine, chlorine and other halogens, methyl, ethyl and other alkyl groups. Phenyl groups may also be appropriate. Preferably substituents at R^1 and R^2 should be stable during synthesis of the marker, during incorporation of the mass label into an oligonucleotide in an automated synthesiser and under mass spectrometry. A wide variety of groups have these properties and may be incorporated into the linker at these positions. It may also be desirable in some embodiments to choose substituents which change the solubility of the linker and alter the rigidity of the linker.

Preferably a covalent linkage is formed in attaching the analyte and/or the reporter group to the cleavable linker. The covalent linkage is not particularly limited provided that the analyte and/or reporter group can readily be attached to the cleavable linker using reactive functionalities attached to the linker and the analyte.

Table 1 below lists some reactive functionalities that may be reacted together to generate a covalent linkage between two entities. Any of the functionalities listed below could be used to form the compounds used in the present invention to permit the linker to be attached to an analyte (such as a nucleic acid or protein) for detection (e.g. by mass spectrometry). If desired, a reactive functionality can be used to introduce a further linking group with a further reactive functionality.

Table 1

Functionality 1	Functionality 2	Resultant Covalent Linkage
$-NH_2$	-COOH	-CO-NH-
$-NH_2$	-NCO	-NH-CO-NH-
$-NH_2$	-NCS	-NH-CS-NH-
$-NH_2$	-CHO	$-CH_2-NH-$
$-NH_2$	$-SO_2Cl$	$-SO_2-NH-$
$-NH_2$	-CH=CH-	$-NH-CH_2-CH_2-$
-OH	$-OP(NCH(CH_3)_2)_2$	-OP(=O)(O)O-

It should be noted that some of the reactive functionalities above or their resultant covalent linkages might have to be protected prior to introduction into an oligonucleotide synthesiser. Preferably unprotected ether, ester, thioether and thioesters, amine and amide bonds are to be avoided as these are not stable in an oligonucleotide synthesiser. A wide variety of protective groups are known in the art to protect linkages from unwanted side reactions.

A short alkyl linkage is appropriate to link the mass marker to the linker, although a wide variety of linkages are available which can be used to link a mass marker to a linker.

The reporter group used in the present invention is not especially limited and may be any group, provided that it is readily detectable and can se related to an analyte to identify the analyte. Typically, the reporter group is a mass marker, that is detectable by mass spectrometry. Other appropriate reporters include fluorophores, radiolabels, chemiluminescent

labels, and electron capture labels. In the compounds of the present invention, the reporter group comprises a mass marker.

In preferred embodiments of the present invention, mass markers disclosed in PCT/GB98/00127, PCT/GB98/03842, GB 9815166.5 and GB 9826159.7 can be employed. The content of these applications is incorporated by reference. PCT/GB98/00127 and PCT/GB98/03842 disclose poly-ether mass markers which are thermally stable, chemically inert and fragmentation resistant compounds, and which can be substituted with a variety of groups to alter properties such as solubility and charge. These mass markers are also preferred for use in the present invention and the content of this application is incorporated by reference. GB 9826159.7 discloses markers which comprise two components, which may be poly-ethers, which are analysed by selected reaction monitoring. These are particularly preferred mass markers for use in the present invention. GB 9815166.5 discloses mass markers that bind metal ions, which are also preferred markers for use with this invention. The content of this application is incorporated by reference. Reporter groups that can be detected by more than one detection means may also be desirable as with, for example, a fluorescent marker that incorporates a radioisotope in its linker and that is detectable by mass spectrometry and reporters of this kind are refereed to as 'multi-mode reporter' groups. Preferred multi-mode reporter groups are detectable by mass spectrometry.

When the mass marker comprises an oligoether or a polyether, the oligoether or polyether may be a substituted or unsubstituted oligo- or poly-arylether. The oligoether or polyether preferably comprises one or more fluorine atom or methyl group substituents, or one or more ^{2}H or ^{13}C isotopic substituents.

It is further preferred that the mass marker comprises a metal ion-binding moiety. Typically, the metal ion-binding moiety comprises a porphyrin, a crown ether, hexahistidine, or a multidentate ligand. Preferably, the metal ion-binding moiety is a bidentate ligand or is EDTA. The metal ion-binding moiety may be bound to a monovalent, divalent or trivalent metal ion. The metal ion is not especially limited. Preferred metal ions include a transition

metal ion, or a metal ion of group IA, IIA or IIIA of the periodic table. Particularly preferred metal ions are Ni^{2+}, Li^{+}, Na^{+}, K^{+}, Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, or Al^{3+}. The presence of a metal ion on the mass marker increases the sensitivity of detection.

It should be noted that this invention is not limited to the mass markers disclosed in the above applications. Any molecule with the correct features can be used as a mass marker. Desirable features include:

- Easily detachable from DNA
- Fragmentation resistant in mass spectrometer
- Single ion peaks
- Very sensitive detection
- Easily distinguishable from background contamination
- Distinguish from DNA
- Be certain that a mass peak is from a mass label
- Compatible with oligonucleotide synthesiser
- Easy to synthesise in a combinatorial manner to minimise number of chemical steps and the number of reagents necessary to generate large number of labels
- Compatible with existing mass spectrometry instrumentation without requiring physical modification.

Mass labels and their linkers can be attached to a nucleic acid molecule at a number of locations in the nucleic acid. For conventional solid phase synthesisers the 5' hydroxyl of the sugar is the most readily accessible. Other favoured positions for modifications are on the base at the 5' position in the pyrimidines and the 7' and 8' positions in the purines. These would all be appropriate positions to attach a cleavable mass with the linker of this invention.

The 2' position on the sugar is accessible for mass modifications but is more appropriate for small mass modifications that are not to be removed.

The phosphate linkage in natural nucleic acids can be modified to a considerable degree as well, including derivitisation with mass labels.

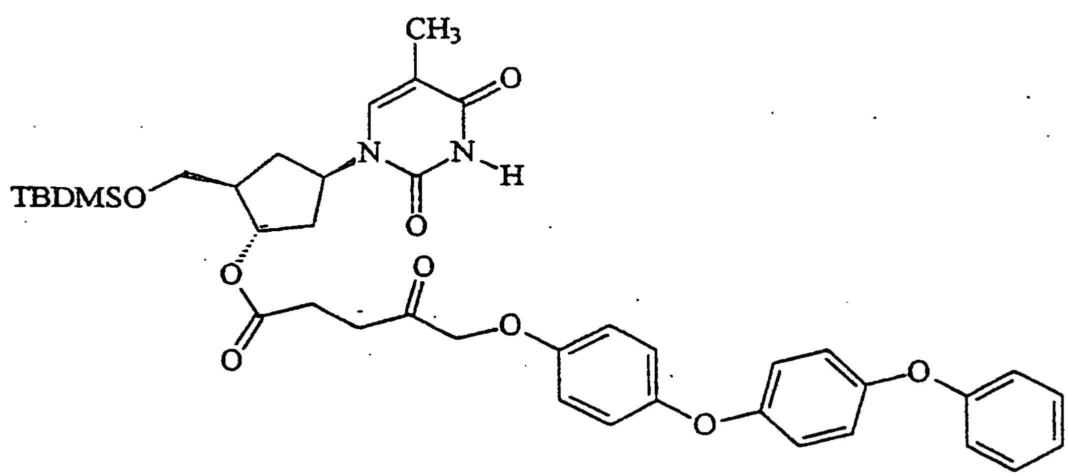
The cleavable linker used in this invention may be cleaved in the ion source of a mass spectrometer by ammonia. However, this invention is not limited to the use of ammonia. Most amines are be capable of separating the mass marker from its cognate oligonucleotide and other nucleophiles may also be used.

The invention will now be described in further detail by way of example only, with reference to the following specific embodiments.

Examples

Synthesis of a mass labelled nucleotide FT23

A thymidine residue with a poly-aryl ether mass label was synthesised as a model compound, shown below:



TBDMS=Tertiary butyl dimethylsilyl

A solution of 5'-O-(tert-butyl dimethylsilyl)-3'-succinyl-thymidine (288 mg, 0.5 mmol) in dichloromethane (3 ml) was treated with three drops of pyridine and then dropwise with a

2 M solution of oxalyl chloride (0.3 ml, 0.6 mmol) in dichloromethane. The reaction mixture was stirred for 90 min at room temp. The solution of the formed acid chloride was added dropwise to an ice-cold solution of (4'-phenoxy)-4-phenoxybenzyl alcohol (146 mg, 0.5 mmol) and pyridine (0.3 ml) in dichloromethane (3 ml). Stirring was continued for 4 h at room temperature. The reaction mixture was diluted with ethyl acetate and washed with a 5 % aqueous solution of $NaHCO_3$ and twice with water. The organic phase was dried with sodium sulphate and the solvent was removed under reduced pressure. The residue was purified by flash chromatography with ethyl acetate/n-hexane (1:1) to yield 73 mg (20 %) of FT 23.

^{1}H NMR ($CDCl_3$): 0.13 (6 H, s); 0.92 (9 H, s); 1.92 (3 H, s); 2.11 (1 H, m); 2.39 (1 H, m); 2.68 (4 H, s); 3.90 (2 H, d); 4.06 (1 H; d); 5.11 (2 H, s); 5.27 (1 H, d); 6.34 (1 H; m); 6.95-7.37 (13 H, m); 7.35 (1 H, d); 8.27 (1 H, br s).

The product of the synthesis was analysed by ESI MS to determine whether the predicted molecular ions were present. The protonated molecular ion was detected at m/z 731 and the molecular less 1 proton was detected at m/z 729. (Mass spectra shown in Figures 4 and 5 respectively)

Mass Spectrometry Analysis of FT23

All data was acquired on a Platform-LC quadrupole instrument (Micromass Ltd, UK) with an electrospray ionisation source.

Figure 2 shows the negative ion mass spectrum of FT23 at 0.35 pmol/µl. This spectrum shows a very clean molecular ion less 1 proton at m/z 729.4.

Figure 3 shows the positive ion mass spectrum of FT23 at 0.35 pmol/µl. This spectrum shows a strong protonated molecular ion peak at m/z 731.5 and a second stronger peak at m/z 753.3 corresponding to a sodium adduct of FT23. Without being bound by theory, it is believed that the poly-ether mass label with the succinate linker is behaving in a similar manner to a crown ether and is binding strongly to sodium to generate these sodium adducts.

Figure 4 shows the negative ion mass spectrum of FT23 at 0.35 pmol/μl in the presence of a PCR product. The molecular ion is not detectable over the background peaks.

Figure 5 shows the positive ion mass spectrum of FT23 at 0.35 pmol/μl in the presence of a PCR product. This spectrum does not show any of the protonated molecular ion peak at m/z 731.4 or the sodium adduct peak at m/z 753.3. Two additional peaks of significance appear in this spectrum. One peak at m/z 739.2 and a second peak at m/z 755.2. These peaks are believed to be end-products of a novel gas phase cleavage reaction discussed below. The two peaks corresponding to gas phase cleavage products are the only major ion peaks in this spectrum.

FT23 is a thymidine derivative which has been protected on the 5' hydroxyl using a tert-butyl dimethylsilyl (TBDMS) protective group. Without being bound by theory, it is believed that in this invention a small quantity of ammonia present in the sample, introduced with the PCR product reacts with the TBDMS protective group according to the mechanism shown in Figure 6. Water appears to react with the TBDMS group to some extent as well. The chemical reactions which take place do not have any charged intermediates. The reaction is visible in the mass spectrometer because of the sodium ion binding behaviour of the mass label used in these studies. The lack of any protonated ions suggest that sodium binding is very strong and that an excess of sodium ions is present which binds to all of the FT23 present. This gives rise to sodium adducts which have a single positive charge. Nucleophilic attack by ammonia generates a trigonal pyramidal intermediate which is energetically unstable. This intermediate rearranges itself with the loss of formyl amine to generate an ion with m/z 739. Similarly nucleophilic attack by water generates a trigonal pyramidal intermediate which is energetically unstable. This intermediate rearranges itself with the loss of methane to generate an ion with m/z 755.

These cleavage product peaks are seen as the majority ions when 0.35 pmol/μl of FT23 is analysed in the positive ion mode in the presence of a PCR product (Figure 5) but are not

observed when the pure compound is analysed (Figure 3). At higher concentrations of FT23 the singly protonated molecular ion peaks are still detectable in the presence of nucleic acid (not shown). In the negative ion mode the pure compound is readily detected at 0.35 pmol/μl but is not detectable at the same concentration in the presence of a PCR product. Ammonia was present in the buffers of the PCR product. The spectra are interpretable if the low levels of ammonia are cleaving the silyl protective group. The levels of ammonia are limiting - at higher concentrations of FT23 the reaction does not go to completion (not shown) but at the lower concentration of FT23 there is sufficient ammonia to completely cleave the molecular ion.

Figure 1 shows how this chemistry can be adapted for use as a gas phase cleavable linker as discussed above.